INTRODUCTION

João Branquinho

The present volume brings together thirteen new essays dealing with a wide variety of important topics in the foundations of the fascinating multi-disciplinary field of studies currently known as Cognitive Science. The purpose of this Introduction is to provide the reader with a synoptic view of the territory covered by the book, especially the main issues and problems addressed, and to give an outline of the central contributions made by each essay to their discussion. It is expected that this will help readers, particularly those less familiar with the area, to be able to discern a background of shared theoretical concerns and general assumptions behind the diversity of topics and approaches displayed by the contributed essays. Given the controversial nature of most issues in the foundations of cognitive science, it could hardly be expected from a description of the territory that it be theoretically neutral; however, we have tried as much as possible to stay close to a set of methodological claims that are very often seen as consensual.

The multi-disciplinary character of the area is clearly reflected in the volume, as witnessed by the fact that a large number of the academic disciplines usually regarded as engaged in the enterprise of cognitive science are represented therein. Indeed, even though some of the essays (e.g. chapter 5) are somehow hybrid and could thus be seen as falling within more than one discipline, a natural way of sorting them out in that respect is as follows: neuroscience (chapters 4, 5, 6 and 7), linguistics (chapter 10), philosophy (chapters 1, 2, 8, 9, 11 and 13), and psychology (chapters 3 and 12). Of course, the relative predominance of philosophy on this list stems from the relative predominance of matters eminently foundational throughout the book, matters having to

do with general questions concerning the relations that hold among the principal protagonists involved the story about human cognition: mind, brain, language, world, and action.

It is worth mentioning at this point that, besides sharing a subject (broadly conceived), a common feature of the essays included in this volume is given in the fact that early versions of them were presented at the Lisbon International Conference on the Foundations of Cognitive Science at the End of the Century. Most people who had the chance to attend the Lisbon meeting in May 1998 would very likely refer to it as a memorable event. It would be thus described not only on the basis of the remembered friendliness of the environment and other traits of the same kind, but mostly on the basis of a feeling that something rather like a "meeting of minds" happened there on the occasion among the several researchers contributing to the present collection. Indeed, the live and cooperative discussion and exchange of views that marked most of the conference sessions, as well as the genuinely interdisciplinary dimension of the debates, very often generated real insight into some of the most interesting and hard issues in the foundations of cognitive science. Although some of these aspects are extremely hard to capture in print, it is not unreasonable to think that what happened in the meeting has had some sort of benign bearing upon the final versions of the essays here included.

THE MIND AS A PROCESSOR OF INFORMATION

How should one characterize the task of cognitive science in a way that would enable us to obtain an integrated picture of the disparate contributions in this volume, a picture on which they are all part of a single theoretical enterprise?

A common and convenient way to do it is to define cognitive science as the scientific study of the mind and of its role in the production of intelligent (purposeful, goal-oriented) behaviour.

Although there is some controversy surrounding the issue, one should perhaps note that the term 'mind' is usually taken in this context to refer not only to the human mind, but in general to the mind of any intelligent information-processing system. Hence, even machines and artefacts of certain kinds – not to speak of animals – should not be ruled out, at least from the outset, as lacking mental activity in the relevant sense. As long as they can be reasonably counted as intelligent information-processing systems, they are assumed to be endowed with minds. (Whether there really are, or could really be, machines that were capable of meeting that condition is in itself a moot issue in the foundations of cognitive science, one that need not bother us here.)

The leading idea underlying the above identification of the subject matter of cognitive science is that minds are basically processors of information; or, given that cognition just is (in one sense) information processing, it is the equivalent idea that minds are essentially cognitive devices. This idea seems to be rather pervasive in contemporary cognitive science. In fact, it even seems to constitute one of the few substantive foundational assumptions in the area that has been relatively immune to dispute so far. In particular, the assumption (or certain versions of it) is clearly endorsed both on classical or symbolic approaches to the mind and on connectionist or non-symbolic approaches, these being the two main opposing methodological schools of thought currently available in cognitive science. Indeed, virtually everyone working in the area would agree in identifying the processes to be studied in cognitive science as being those commonly involved in the information-processing activity of the mind (or of the brain), namely the processes of receiving, storing, retrieving, modifying, and

transmitting information of various kinds. Accordingly, virtually everyone working in the area would be prepared to count as paradigmatic instances of information processing everyday phenomena such as visual perception and language comprehension, some important aspects of which are dealt with in chapters 12 and 7 of the present collection (respectively).

Thus conceived, the field of cognitive science covers not only cognition proper, the usual paradigms of which are mental events and processes such as conscious thoughts and inferences and intentional mental states like beliefs and desires, but also any other mental phenomena in which information processing happens to play a central role. Hence, insofar as they can be subsumed as instances of information-processing activity (which in many cases is clearly the case), a variety of mental states and events traditionally grouped under the heading of 'experience' and often contrasted *in limine* with cognitive states and events also fall within the scope of cognitive science; among those states and events are notoriously sensations and perceptions, for instance visual experiences such as the experience of seeing a red object moving around in one's visual field and auditory experiences such as the experience of hearing a piano sonata. In other words, both propositional attitudes – as philosophers call psychological states like belief and desire – and sensory experiences are part of the subject matter of cognitive science (note that on some views cognitive science should be interested in propositional attitudes just to be able to eliminate them from proper scientific inquiry). Yet, only states of the former kind are strictly cognitive in the following sense. In order to be in one of those states a system must possess and employ an appropriate set of concepts or ways of categorizing things; for instance, a system or organism cannot be in the state of believing that a cow is in front of it without having the concept *cow*, but it certainly can see a cow in front of it (a visual experience) without having such a concept.

So, one way to form a unified picture of the plurality exhibited by the essays gathered in this volume is to see them all as engaged in the task of explaining some aspect or other of the workings of the mind as an information-processing device. (As we shall see in a moment, this description is only partially accurate and needs qualification; but it will do for immediate purposes.)

REPRESENTATION AND COMPUTATION

The widely held assumption that the mind is basically an information-processing device is normally supplemented by two other relatively uncontroversial foundational claims. These claims complement that assumption by providing further specifications of what is going on in the information-processing activity. In particular, they can be seen as providing answers to the following two questions. (a) What is in general processed by the cognising mind (or brain)? In other words: what is the nature of the information manipulated by the mind (or brain)? (b) What is the general form of the processing? What is it like and how does it proceed?

The first claim answers question (a) by specifying the objects over which the information-processing activity is defined as being essentially mental representations. Roughly, these are items in the mind or brain of a given system that in some sense "mirror", or are mapped onto, other items or sets of items; the latter are typically items external to the system: objects, events, situations, etc, in the world (typically, in the immediate environment of the system). Thus, it is assumed that cognition operates on mental representations, in the sense that these are basically what is being processed by the minds or brains of intelligent information-processing systems when they are performing given cognitive tasks; these tasks range from relatively simple ones, such as

detecting the presence of a predator in the surroundings, to very complex ones, such as proving a mathematical theorem.

Note that mental representations might come in a wide variety of forms, there being no commitment in the claim itself to a specific kind of representation or to a particular sort of representational vehicle. According to taste or theory or purpose, mental representations might be thought of as images, schemas, symbols, models, icons, sentences, maps, and so on. Their job is to provide systems with the information they need to control their behaviour and guide their interactions with the environment. Mental representations are thus supposed to depict, by means of any vehicle that turns out to be appropriate to do that job, not only aspects and states of the outer world, but also aspects and states of the inner world, namely internal states of the system.

The second claim answers question (b) above by identifying the nature of the processing, by describing in general what goes on inside the "black box" (to use a familiar metaphor). The processing is characterized as being essentially computational, as consisting of a series of computations executed by the system (or by its mind or brain). Given the first claim, it follows that the computations involved in cognition are defined over a set of mental representations. Roughly, this means that the cognitive tasks performed by the mind or brain characteristically consist in generating in an effective way certain mental representations as outputs on the basis of certain mental representations given as inputs. The operations performed by the visual system or by the language faculty in humans are clear cases of computations in this sense: in the former case they take 2D patterns (retinal images) as inputs and yield representations of 3D scenes as outputs; in the latter case they take acoustic representations of utterances as inputs and yield semantic representations as outputs.

Now it might turn out that even these two foundational claims, as well as the underlying conception of the cognising mind as a processor of information, come to be challenged at some point. Yet, as often emphasized in the literature in the area, they have so far enjoyed the special status of shared principles or background assumptions. In other words, they have been regarded as defining the whole enterprise of cognitive science. It comes then as no surprise that both claims are endorsed – even if tacitly in most cases – throughout the present collection, at least if one relies on appropriately generic formulations as the ones provided above. Of special interest in this context is the fact that the claims in question seem to cut across the methodological dispute between symbolic and non-symbolic models of cognition, a dispute that comes to the surface at several points along this volume (see below).

Anyway, an additional way of forming a unified picture of the variety of contributions and approaches contained in this book is to see them as somehow presupposing those two assumptions about the nature of the cognitive workings of the mind: the assumption that the mind represents and the assumption that the mind computes.

Of course, the controversy installs itself immediately once one tries to flesh out those foundational claims and be more specific both about the nature of the posited mental representations and about the nature of the computations defined over them.

Thus, one might want to hold the view that the mental representations involved in cognition are intrinsically symbolic or linguistic; that they are, to take a familiar proposal, sentences in a "language of thought". And one might want to couple this idea with the sister idea that the computations involved in cognition are like the operations characteristically performed by digital computers, that they consist in the application of purely formal rules or procedures to syntactically defined items (language-like mental

representations). These views represent, however, specific methodological proposals, particular ways of fleshing out the above claims, and are thus far from being immune to criticism. Actually, their joint endorsement partially defines the classical or orthodox approach to cognitive science, a view sometimes known as the Representational Theory of Mind. Despite being still very influential, the classical approach is highly controversial. As far as the present volume is concerned, it is explicitly assumed in James Higginbotham's essay (chapter 10) with respect to language understanding and endorsed in Zenon Pylyshyn's essay (chapter 12) with respect to visual representation and object tracking; but it is explicitly rejected in the essay by Ilya Farber, Will Peterman, and Patricia Churchland (chapter 4) with respect to spatial representation and knowledge.

Another well-known way of fleshing out those claims is given in the rival connectionist picture of the workings of the mind. On this view there is still room for mental representations, as well as for computations performed on them. But connectionist representations and computations are of a radically different ilk. For one thing, mental representations are here definitely non-symbolic and non-linguistic: they do not represent in the way that words and sentences represent. According to usual versions of connectionism, the mind's information-processing activity should be regarded as being intrinsically linked to, and impossible to divorce from, its physical medium: the brain. Thus, the computational processes involved in cognition are highly sensitive to the physical system in which they are implemented and cannot be assimilated to the purely formal operations performed by digital computers. Rather, connectionist computations proceed through the simultaneous operation of a large number of basic processing devices – the so-called units, nodes, or artificial neurons (the counterparts of the brain's neurons) – that interact with one another and form a

complex multilayered network. In contrast with standard classical or symbolic computations, which are both local and sequential, connectionist computations are distributed and parallel (at least on some branches of connectionism); semantic content and representational properties might accordingly be assigned simultaneously to several nodes, and so connectionist mental representations are also distributed and not local. Cognition is modelled on the basis of artificial neural networks, structures that mimic in a simplified way the complex operations and processes of the brain. In spite of being highly controversial too, the connectionist picture has been embraced or explored by many cognitive scientists, especially those working in neuroscience and cognitive neurobiology. As far as the present volume is concerned, the approach is explicitly endorsed in the paper by Farber, Peterman, and Patricia Churchland (chapter 4) and in Paul Churchland´s essay (chapter 5), where cognitive skills like those involved in spatial representation and moral discrimination (respectively) are modelled by means of neural networks and are given a connectionist treatment.

BEYOND COGNITION?

Notwithstanding what has been said so far, recent developments seem to indicate that, after all, cognition is not all there is to the mind that is of interest to cognitive science. This seems to be the case even if one adopts a liberal view about the scope of cognition, a view on which it includes not only the realm of strictly cognitive or "conceptual" events and states, events and states endowed with propositional content, but also the realm of experience (sensation and perception).

The above remark has a certain air of paradox, since the modifier 'cognitive' employed in the term 'cognitive science' suggests a restriction to cognition, to the study of information-processing activity as it occurs in perception, memory, knowledge,

language understanding, reasoning, problem solving, and so forth; that is to say, it suggests a restriction to the cognitive side of the mind. Nevertheless, the fact is that a gradually increasing focus has been lately placed within cognitive science upon what prima facie are non-cognitive mental processes, states and events, especially those belonging in the so-called volitional and emotional departments of the mind. (We employ here for convenience the classical tripartite picture of mental life, a picture on which the mind is seen as divided into the separate areas of cognition, conation, and affect.) The paradox is only apparent, of course. For either the mental phenomena in question are in the end correctly describable as genuine instances of information-processing activity and turn out to be at bottom cognitive in nature; or an adequate redescription of the province of cognitive science is needed, a redescription which should inter alia include a revision of the conception of the mind as a processor of information.

At any rate, a considerable amount of research has been going on in cognitive science into the nature of a range of mental phenomena about which it is not straightforward (to say the least) that they can be reduced to, or that they centrally involve, cognition or information processing.

Emotion is a case in point. Emotional processes have been in recent times at the centre of a great deal of work in cognitive science, work carried out not just in those disciplines where there is a traditional concern with such phenomena – disciplines such as psychology and philosophy – but also in other disciplines, most notably cognitive neurobiology. One of the hot issues here is precisely the extent to which cognition is involved in emotion. Roughly, the debate in the area opposes cognitivist accounts, on which emotion is at bottom a species of cognition, and non-cognitivist accounts, on which emotion cannot be reduced to cognition. That cognition is present in emotion

seems to be beyond doubt; as shown by Antonio Damasio in his contributed essay (chapter 6), emotional processes invariably contain a clearly cognitive ingredient in the form of an evaluative judgement made by the subject. The problem is rather whether or not cognition, in the sense of information processing, is the essence (so as to speak) of emotion. One should note that the converse issue, the issue about the extent to which emotion is involved in cognition, is also a key issue in the area; it is extensively discussed by Antonio Damasio in his book *Descartes's Error* (A. R. Damasio 1994).

Consciousness, as represented in a range of higher-level mental states traditionally viewed as mysterious in the sense of irreducible to objective or scientific description and explanation, is also a case in point. Of special interest in this context is the form of consciousness Ned Block calls *phenomenal* consciousness (see chapter 1 of this volume), and in particular the set of sometimes associated properties known in the philosophical literature as *qualia*. These are the subjective qualities of sensory experience, vivid examples being the allegedly unique phenomenological aspects of the experience of having an orgasm and the allegedly unique phenomenological aspects of the experience of listening to the sound of Galician *gaitas* (bagpipes). It is also a matter of intense dispute the extent to which consciousness in general, and phenomenal consciousness in particular, can be given a satisfactory account in terms of the information-processing model.

This broadening of the scope of cognitive science, so as to make room for mental phenomena that are not clearly cognitive in nature (even if they turn out to be so in the end), is nicely represented in the present collection. Emotion and feeling are addressed in Antonio Damasio's essay (chapter 6), consciousness in Block's essay (chapter 1), and morality in Paul Churchland's essay (chapter 5). Furthermore, if the main line of reasoning advanced by John Searle in his essay (chapter 13) is sound, then

modelling human action in terms of information processing, in the sense of construing it as systematically arising out of some appropriate set of desires and beliefs, gives us a wrong picture of the relation between mind (volition) and action.

FOUNDATIONAL ISSUES

Apart from foundational work on methodological issues, in which one should include the discussion of background assumptions and guiding principles such as those considered earlier on, there are in the foundations of cognitive science – just as in the foundations of other scientific subjects – two main interrelated kinds of research. One kind of work is concerned with the clarification and elucidation of the most basic or central concepts employed in the discipline; thus, the task of explaining such concepts as *information*, *knowledge*, *concept*, *consciousness*, *cognition*, and so on, belongs in this segment of foundational work in cognitive science. Another kind of foundational work is concerned with the investigation of a set of highly general and speculative issues, issues about the nature of mind and cognition that are motivated by developments occurring in the several disciplines contributing to cognitive science. These questions are predominantly philosophical in character and some of them, e.g. the so-called mind-body problem, happen to have been discussed by philosophers for centuries. Thus, the age-old inquiry into the nature of consciousness, the discussion around the question whether it is possible to simulate mechanically intelligent behaviour, and the reflection on whether cognition and thought are possible without language, all belong in this segment of foundational work in cognitive science.

One should notice that foundational work of the latter kind ranges from research that is straightforwardly philosophical in nature, in which general questions of that sort are explicitly addressed, to research that, although not dealing explicitly with such

problems, generates results that are highly relevant to their discussion. A related remark is that, by analogy with foundational work in other scientific disciplines, it is plausible to regard foundational work in cognitive science as being in most cases just a natural continuation of non-foundational work, the differences between the latter and the former being essentially differences in degree of generality and scope and not differences in kind. Accordingly, even though for obvious reasons philosophy should be seen as playing a pivotal role in the foundations of cognitive science, it is by no means accurate to view it as having the monopoly over research in the area. Indeed, it becomes apparent from the previous observations that important foundational work of any of the kinds mentioned above can be (and has been) carried out in any one of the other academic disciplines involved in cognitive science. A clear illustration can be found in the essays contained in the present collection. One can find conceptual work in the foundations of cognitive science (foundational work of the first kind) not only in Block's important distinction between two concepts of consciousness, *phenomenal* consciousness and *access*-consciousness (chapter 1), but also in Antonio Damasio's important clarification of the notions of emotion and feeling (chapter 6); likewise, one can find "philosophical" work in the foundations of cognitive science (foundational work of the second kind) not only in Donald Davidson's defence of a broadly linguistic approach to thought and cognition (chapter 8), but also in Susan Carey's discussion of certain forms of cognition in infants and nonhuman primates and in her conclusion that they are radically non-symbolic and non-linguistic (chapter 3).

Three research themes have acquired a salient position in foundational work of the second kind (of course, there are natural intersections with foundational work of the first kind and with foundational work on methodological issues). They all relate to the connections holding between the mind, especially the human mind, and other

intervening elements in any adequate account of its workings as a cognitive device. A first aggregate of topics concerns the relations between the mind and the brain – or, to be precise, the relations between the mind and the central nervous system. A second theme concentrates on the relations between mind and language, where by 'language' it is meant either a natural language like English or any other symbolic system of representation. Finally, a third group of issues concerns the relations between mind and world, including aspects about the evolutionary history of both and also the connections between the mind and intentional behaviour and action upon the world.

These three research themes provide us with a convenient way of sorting out the principal contributions made by each essay in the present collection to the foundations of cognitive science.

MIND AND BRAIN

A first group of essays – chapters 1, 4, 5, 6, and 7 – address mostly issues concerning the relations between mind and brain.

It is widely known that mental activity has some sort of physical implementation in the brain. Indeed, there is good evidence that mental states, events, and processes are directly correlated with certain states, events, and processes in the brain – or, in general, in the central nervous system – in the sense that they co-vary with the latter in regular and predictable ways. The search for neural correlates (as they have been called) for a wide range of mental phenomena has always been accorded a key position within cognitive science, especially within those disciplines – the neurosciences – that are particularly suited to the task. More recently, the focus has somehow been shifting from a concern with cognitive phenomena that are more or less familiar in that respect, like

language processing or memory, to a concern with prima facie non-cognitive and much less familiar aspects of mental life, like consciousness and emotion.

One should also note that the inquiry into the neural basis of mental states, events, and processes is regarded not only as being intrinsically valuable but also, on some non-symbolic views of mentality and cognition, as disclosing the nature or essence of those states, events, and processes themselves. Those views assume, roughly, that the physical implementation of mental phenomena in certain systems in the brain is something that defines them (at least partially); they contrast thus with symbolic views, on which the neural realization of mental phenomena – the "hardware" on which they are actually "run" – is in general seen as largely irrelevant to their identity.

In his contribution to this collection (chapter 1) Ned Block discusses some aspects of the long-standing issue about the neural basis of consciousness. He begins by noting that, as employed by philosophers and cognitive scientists, the term 'consciousness' is ambiguous as it is used to express different concepts of consciousness: phenomenal consciousness and acess-consciousness. The former is identified with experience: one is phenomenally conscious if one is having an experience; sensations are typical cases of conscious states of this kind. The latter is defined thus: "a representation is acess-conscious if it is actively poised for direct control of reasoning, reporting, and action" (p. 3); thoughts and propositional attitudes are typical examples of conscious states of this kind. The case is then made for the distinctness of the concepts by showing that it would be possible for access-consciousness to be instantiated without phenomenal consciousness being thereby instantiated and also, more controversially, that the latter might occur without the former. Yet, as one can see from pairs of concepts such as *water* and *H20*, distinct concepts may nevertheless be co-referential; so nothing would prevent phenomenal

consciousness and access-consciousness from being actually correlated with the same system in the brain.

Armed with this distinction Block goes on to examine two views, which he diagnoses as guilty of conflating the two notions of consciousness. One is John Searle's account of consciousness as not being necessarily involved in habitual, routine, and memorized activities such as e.g. driving a car. The other is Francis Crick and Christof Kock's claim that the area in the brain associated with vision and known as V1 is not part of the neural correlate of consciousness. Block contends that their argument to that claim as conclusion faces the following dilemma. If 'consciousness' means access-consciousness, then the argument is trivial; if phenomenal consciousness is meant, then it is unsound. Either way, the argument does not succeed in establishing the substantive conclusion it aims to establish. However, two positive claims are added in Block's essay to this negative upshot: one is the claim that Crick and Koch's contention that V1 is not part of the neural correlate of consciousness may in the end be right but on the basis of independent considerations, considerations that are nevertheless implicitly contained in their own work; the other is the claim, also extractable from considerations made by Crick and Koch, that the two concepts of consciousness may not be co-referential as they may have different (though overlapping) neural correlates.

The nature of the processes involved in emotion and feeling and their neurobiological basis are the topic of Antonio Damasio's essay in this volume (chapter 6). Antonio Damasio is in general sympathetic to William James's approach to emotion, especially to James's idea that the essence of the emotional process is a "sense of the body", an emotion being basically a perception of a set of physiological changes that take place in the body in response to a given situation or event. Nevertheless, James's account should be subjected to two important kinds of amendment. The first concerns

the role of cognition in emotion. According to Antonio Damasio, James underestimates in general that role. In particular, in James's account no place is assigned to the clearly cognitive and non-automatic evaluative judgements that precede in most cases the emotional responses. Antonio Damasio argues that a mental appraisal of the significance of the stimuli which cause an emotion is in many cases a crucial component of the emotional process; it is notably present in most emotions we experience as adults. The second correction to James's account concerns the role of the body in emotion. Antonio Damasio's claim is that such a role is in a sense overestimated by James. In particular, there is no room in James's account for a neural mechanism that would be able to generate, without the intervention of the body (or else in a supplementary way), that state of awareness of body changes which is the distinctive mark of an emotion. Antonio Damasio argues that evidence clearly shows that such a mechanism is available, that the brain has in itself the resources to form an "as-if-body" state (p. 106), a state that depicts the body as if the body were being activated and modified in the way characteristic to emotion. Hence, contrary to James's idea, the interposition of the body in the emotional process is not necessary (even though it is a fact in many cases).

Within the highly complex phenomenon of emotion Antonio Damasio makes a subtle distinction between (a) emotion proper, which he identifies with the expressive element, the public manifestation of the phenomenon, and (b) feeling, which he identifies with the internal, subjective experience. The core of an emotion is given in a set of dispositions to respond that induce a collection of changes both in the state of the body and in the state of the brain, these emotional responses being preceded by the mental appraisal process mentioned above. The subject's representation of those body and brain states, which assumes the form of a complex set of mental images, is what

constitutes the experiential component, the feeling. According to Antonio Damasio, feelings are thus a privileged means the brain has of providing us with knowledge of our body, not in the everyday sense of the term, but in the sense of "cognition of our internal milieu, visceral, and musculoskeletal states"; feelings are nicely described by him as "the first step in letting us mind the body" (p. 107). The presence of both non-cognitive (or automatic) and cognitive elements in the emotional process is paralleled at the level of the brain systems that constitute the neural basis of emotion. Thus, the traditional picture of emotion as being strictly connected with the limbic system and older brain structures, especially the amygdala and the anterior cingulate cortex, is replaced in Antonio Damasio's account with a richer picture that also makes room for brain systems that support the *cognitive* aspects of the emotional process; on this picture modern brain structures, in particular the neocortex, are seen as playing a key role.

On Antonio Damasio's view emotion is not an excrescence, a redundant feature of mentality; it has a salient regulatory function and contributes in important ways to the overall success of the interactions between organisms and their environments. Another feature of the mind that is surely no less crucial in that respect, having a conspicuous "survival value", is spatial representation and reasoning. Indeed, a great deal of successful behaviour, behaviour that satisfies our needs and wants, is highly dependent upon our ability to represent space, to perceive the relative positions of independently existing things with respect to our body (or head) and to one another. This ability is involved in the execution of both very modest tasks, such as reaching for an object and grabbing it, and rather sophisticated ones, such as solving a geometry problem. As one can see from simple cases where the ability is exercised, forming an internal representation of space requires in general the joint use of sensory and motor information, as well as the integration of various sensory modalities (sight, touch, etc.).

The nature and structure of spatial representation and reasoning, and how it relates to other forms of mental representation, have become a topic of great interest in the foundations of cognitive science.

In their contribution to this volume (chapter 4), Ilya Farber, Will Peterman, and Patricia Churchland focus on the issue of how brains and nervous systems, especially mammalian brains and nervous systems, are able to represent space and carry out spatial reasoning. Their central thesis is that spatial representation, as it occurs in humans and other mammals, is fundamentally nonsymbolic. This means that the way in which it happens to be physically implemented in the brain and nervous system, the specific configuration of the underlying neural systems, is a constitutive feature of our ability to represent space, something that determines the intrinsic nature of spatial representation. Given *inter alia* the important position this form of representation occupies in human cognition (for example, it is clearly involved in forms of self-representation such as the representation of one's own body), that thesis is intended as part of a broader picture of mental representation as nonsymbolic in the above sense. Accordingly, Farber, Peterman, and Churchland mount a vigorous attack on symbolic or linguistic approaches to cognition in general. Their main target is the familiar brand of symbolic approach known as functionalism. This is a view on which the nature of mental states can and should be described without any reference to the physical structures in which they are implemented; they are to be individuated rather in terms of their function within a given network of mental states.

In support of their nonsymbolic approach to spatial representation Farber, Peterman, and Churchland draw on three kinds of empirical evidence: (i) behavioural data from animal psychology; (ii) data from neuroscience about the neural basis of spatial representation in mammals; and (iii) data from lesion studies, particularly studies

of impairments of spatial representation and reasoning that result from damage to the parietal lobe. They argue that the symbolic view is in real trouble when confronted with data coming from any of these sources. As to data of kind (i), Farber, Peterman, and Churchland take them as indicating that spatial reasoning based on representations of object-centred space is carried out in organisms concerning which it seems absurd to postulate any relevant representational systems that are language-like or symbolic. Baboons, nutcrackers, bears, ravens, and rats are able to solve a host of problems – the hiding problem, the trapeze problem, the problem of finding or retrieving food, etc. – that require complex spatial reasoning, reasoning surely not different in kind from that which is carried out by humans. In particular, in order to solve such problems animals must be able to represent accurately the relation of their own bodies to various objects and other bodies in space. The proponent of the symbolic view is then confronted with the following dilemma: either she claims that there is indeed a difference in kind between human and animal representation of space, which is implausible; or they rule out spatial representation in general as being nonsymbolic, in which case they would have to find a way to discriminate between it and other cognitive skills of the same level of complexity. As to data of kind (ii), they are taken by Farber, Peterman, and Churchland to show that the mammalian brain and nervous system is able to build representations of object-centred space in accordance with a radically nonsymbolic model. The region of the brain that is responsible for producing "objective" representations of space is the posterior parietal cortex and the crucial explanatory hypothesis invoked by Farber, Peterman, and Churchland in that respect is the Pouget-Sejnowsky Hypothesis. According to this hypothesis, the posterior parietal cortex generates *basis functions*, which compute such representations; a basis function is the product of two kinds of processing units (neurons), eye position units and retinal

position units, such a product being computed by hidden units (interneurons in area 7 of the posterior parietal). According to Farber, Peterman, and Churchland, these results from neuroscience provide us with strong evidence against symbolic models of spatial representation. Finally, as to data of kind (iii), they claim that impairments in spatial reasoning such as hemineglect, the tendency to ignore or neglect objects in particular regions of space, are hardly accountable in terms of symbolic approaches to spatial representation (whereas nonsymbolic approaches have no difficulty in explaining them).

Language processing, the way the human mind is able to produce and understand spoken or written language, is a research area in cognitive science that has many significant implications for foundational inquiry into the nature of the cognising mind and its relation to the brain. The neural basis of some central aspects of language processing is the issue addressed by Hanna Damasio in her essay (chapter 7). From the perspective of neuroscience and working at the level of large-scale systems in the human brain, she deals with two sorts of cognitive tasks that are crucially involved in language production and comprehension: word retrieval and concept retrieval. The former consists in the ability competent speakers of English (say) normally exercise when, given a perceptually presented item (a particular person, an animal of a certain kind, an artefact of a certain kind), they recognize it by coming up with a correct word for it or for the kind it belongs to (a proper name like 'John Smith', a common noun like 'cat', or a common noun like 'hammer'). The latter consists in the ability we exercise when, given a perceptually presented item (e.g. a cat), we recognize it by describing it in an appropriate way, as having such and such salient features and properties (e.g. 'small feline that people keep at home…'). As Hanna Damasio points out, the fact that the former ability has not been exercised for some reason (one sometimes forgets the words) is not evidence that concept retrieval has thereby failed: one might have a

concept for the presented item, as incorporated in some adequate description one might produce, without being able to produce the name. This is paralleled at the level of the corresponding neural systems: an impairment of the area of the brain that supports the retrieval of names for persons (the left temporal pole) may leave intact the concept-retrieval ability (see p. 110). Of course, word retrieval is taken as being in general sufficient for concept retrieval in the above sense.

Hanna Damasio investigates the neural systems underlying word retrieval and concept retrieval. She draws heavily on studies of two kinds, Lesion studies and PET (Positron Emission Tomography) studies, conducted by her and in which recent sophisticated techniques for correlating cognition and brain are applied. These techniques are: functional imaging, by means of which brain activity is indexed when a subject is performing a certain cognitive task; and a modern version of the lesion method, by means of which hypotheses about the neural basis of cognitive processes are tested by considering damaged areas of the brain. It turns out that the results obtained in the Lesion studies about the neural systems involved in word and concept retrieval largely coincide with the results obtained in the PET studies. The general conclusions Hanna Damasio draws from these studies are as follows. First, the traditional anatomical map of language areas, or neural correlates of language processing, comes out as rather incomplete. According to such a map, language correlates in the brain are mostly restricted to two well-known areas (both located in the left hemisphere): Broca's area and Wernicke's area. Hanna Damasio argues that this picture is very simple and should be replaced by a more complex one, a picture on which many other regions of the brain (some of which in the right hemisphere) are involved in language production and comprehension; these regions are connected by bi-directional pathways and form a complex system. For example, the retrieval of names for particular persons seems to

correlate with the left temporal pole, an area that is distant from both Broca's area and Wernicke's area. Furthermore, the neural correlates for the retrieval of concepts for persons, on the one hand, and for the retrieval of concepts for tools, on the other, seem to be situated in different hemispheres. Second, the neural correlates for word retrieval with respect to a given item are separable from the neural correlates for concept retrieval with respect to the item in question; word retrieval seems to correlate with regions in higher-order cortices of the left temporal lobe, whereas concept retrieval seems to correlate with regions in "higher-order cortices in right temporal polar and mesial occipital/ventral temporal regions, and in lateral occipital-temporal-parietal regions" (p. 116-7). Third, if we focus our attention just on word retrieval, we see that neural correlates seem to vary according to the conceptual category to which the presented item belongs. The systems in the brain that support word retrieval with respect to persons, word retrieval with respect to animals, and word retrieval with respect to tools seem to be partially segregated from one another; the first systems appear to be located in the left temporal pole, the second systems in the left infero-temporal cortex, and the third systems in the posterolateral infero-temporal cortex. Similar results apply to concept retrieval, which seems also to vary according to the conceptual category of the presented object.

In his contribution to this collection (chapter 5) Paul Churchland deals with the issue of how the human brain and nervous system is able to produce such things as e.g. moral virtues (as well as vices, of course), how the brain is able to generate higher-level mental functions such as those associated with moral representation and knowledge. Approaching the issue from a connectionist standpoint, Paul Churchland's central contention, for which he provides detailed argument and evidence, is that a vast number of significant phenomena involving morality can be given an integrated explanation in

terms of neural-network theory. The basic idea is that the concepts and methods of connectionist cognitive modelling, which are employed to account for our performance of a host of natural cognitive tasks (such as visual perception or language processing), can be successfully applied to the moral realm too. Connectionist cognitive modelling proceeds by using simulated networks of simple processing units to explain human cognition. Such units resemble in a number of relevant respects individual neurons in the human brain, so that we obtain thus a model of the cognitive workings of the brain. Of course, the model is oversimplified: for one thing, neurons in living brains clearly outnumber processing units in artificial networks. But it gives us an overall correct picture of cognitive activity as consisting in complex interactions among large numbers of simple processing units.

Paul Churchland focuses on issues in the branch of moral theory known as Metaethics. This is *inter alia* the study of moral cognition, of the nature of moral judgement and moral knowledge, and it includes questions about how these are acquired and exercised. A cardinal assumption adopted in the essay, an assumption needed to warrant the extension of neural-network theory to the moral realm, is the view that moral cognition is essentially a set of skills, a set of complex perceptual, reflective, and behavioural skills that a morally knowledgeable adult possesses (p. 79). As such it differs only in detail from any other form of human cognition, most notably scientific knowledge. (One should note at this point that the parallel between moral cognition and progress and scientific cognition and progress plays an important role in the main argument of Paul Churchland's paper.) Hence, the formation and deployment of the set of skills embodying moral knowledge and representation should be, in principle, as capable of being modelled by means of neural networks as those skills embodying any other form of knowledge and representation.

Paul Churchland begins by highlighting the merits of a prospective research program based upon a systematic collaboration between the moral discipline of metaethics, on the one hand, and cognitive neurobiology, on the other hand. Such collaboration assumes the form of a bilateral interaction, with results from research undertaken at the micro structural level contemplated by neural-network theory simultaneously feeding and being fed by high-level reflection in the domain of moral knowledge. The program is motivated by an analogy established with the interaction existing between cognitive neurobiology and other philosophical disciplines, especially the philosophy of science, an interaction Paul Churchland sees as theoretically fruitful and insightful. Then he goes on to examine in detail, from the point of view of neural-network theory, virtually all of the topics and issues commonly addressed in metaethics: moral knowledge, moral learning, moral perception, moral ambiguity, moral conflict, moral argument, moral virtues, moral character, moral pathology, moral correction, moral diversity, moral progress, moral realism, and moral unification. Moral knowledge, for example, is described (p. 81) in neural-network terms as being embodied in an intricate configuration of weighted synaptic connections; such connections partition an abstract space of possible activation patterns of neurons or processing units into a hierarchical set of prototypical moral categories ("morally significant action"/"morally nonsignificant action", "morally good action"/"morally bad action", etc.).

Paul Churchland sees the general account that emerges from such an application of connectionist cognitive modelling to the above moral phenomena as belonging in a particular pre-existent tradition or school of thought in metaethics, a tradition known as Virtue Ethics. Virtue Ethics is a family of views about the nature of morality and moral judgement that go back to Aristotle and have been recently endorsed by ethicists such as

Mark Johnson, Owen Flanagan, and Alasdair MacIntyre. The leading idea of Virtue Ethics is that morality is to be fundamentally explained, not in terms of some fixed set of ultimate principles or rules that would govern moral judgement, but in terms of inner characteristics or virtues that individuals gradually acquire on the basis of complex dealings with the social environment. According to Paul Churchland, the account of morality provided by cognitive neurobiology fits nicely with Virtue Ethics. However, in spite of recognizing a large area of agreement, he closes his essay with a critical examination of the treatment given to the issue of moral progress by Virtue Ethics theorists Flanagan and MacIntyre. Both these writers endorse, for different reasons, a frankly sceptical view with respect to moral progress, claiming that there are no real advances in the sphere of human moral consciousness. Paul Churchland argues that such a scepticism is misplaced by drawing on the already noted parallel between moral cognition and scientific cognition, a parallel supported by the findings of neural-network theory; since moral knowledge does no differ in kind from scientific knowledge, to the extent that there is genuine progress in the domain of science, to that very extent there is also genuine progress in the moral realm.

MIND AND LANGUAGE

A second group of essays – chapters 3, 8, and 10 – can be seen as dealing mainly with issues concerning the relations between mind, especially cognition and thought, and language.

Questions about the mind/language connection, e.g. questions about the extent to which human thought and cognition are constitutively determined by their usual verbal expressions in a natural language, are nowadays best addressed from a multidisciplinary perspective. They have always been of interest to philosophers, occupying a prominent

position in current debates in philosophy of language and in philosophy of mind. Recently they have also attracted the attention of many researchers working in other branches of cognitive science, particularly linguistics, psychology (cognitive and developmental), anthropology, and computer science; a number of significant results obtained in these disciplines have helped shed new light into a number of traditional problems about the nature of the relation between human cognition and thought, on the one hand, and their linguistic clothing, on the other.

The mind/language connection is very often thought of as bi-directional, as consisting in some sort of mutual dependence between its terms. Yet, this way of looking at the connection may naturally give rise to conflicting views. Taken in the mind-language direction, the connection seems straightforward. Language processing, the ability to learn, use and understand a natural language, is in some sense a mental activity. The key concept here is meaning, a concept under investigation in many areas of cognitive science. Indeed, the meaningfulness of otherwise meaningless sequences of sounds or marks is something that the mind somehow imposes upon language. But the specific role assigned to the mind in that respect, as well as the nature of the mental and cognitive machinery involved in language production and comprehension, are matters of intense dispute. On some accounts cognition and thought are claimed to be prior to language either metaphysically or in the order of theoretical explanation. These are different, independent priority claims. The priority of thought in the latter sense usually means that linguistic meaning should be ultimately explained in terms of mental states, complex beliefs and intentions in the minds of speakers and hearers; in the former sense, it usually means that there could be thought without language. Taken in the language-mind direction, the connection is sometimes captured in the claim that cognition and thought are shaped by language in the sense that the availability of

thoughts of certain kinds is dependent upon the availability of certain linguistic resources adequate to express them; in other words, the claim is that there are thoughts we are able to apprehend or express but would not be able to apprehend or express if we lacked a language endowed with such resources. Likewise, on some accounts language is claimed to be prior to cognition and thought either metaphysically or in the order of explanation. The priority of language in the latter sense usually means that our capacity for entertaining thoughts, especially our capacity for having propositional attitudes such as beliefs and wishes, should be ultimately explained in terms of certain relations holding between us and linguistic items such as sentences of a natural language; in the former sense, it usually means that thought is not possible without language.

The question whether language is prior to thought in the metaphysical sense receives an explicit positive answer in Donald Davidson's contribution to this collection (chapter 8) and an implicit negative answer in Susan Carey's contribution (chapter 3). Of course, a genuine theoretical conflict would arise here only if roughly the same thing were meant by the term 'thought' on both views, which is far from being clearly the case.

Carey takes as her starting point a certain set of cognitive primitives, concepts that are basic ingredients in our conceptual scheme or system of representation of the world. She considers two key kinds of basic representational resources: concepts of object and concepts of number. The former include the concept referred to in the psychological literature as the *object* concept, in the sense of our general concept of a *bounded physical object* or *concrete particular*; and also representations of numerical identity and distinctness between objects – one and the same object versus two distinct but physically similar objects – as given in basic quantifiers such as *one*, *another*. The latter include our representations of the series of natural numbers, especially the first

integers 1, 2, 3, and the associated counting system. These concepts are expressed in several ways in the lexicon or syntax of any natural language and play an important role in our overall representation of the world.

The crucial foundational questions Carey raises in her paper about the above set of cognitive primitives are as follows. Are these concepts language-dependent, in the sense of pressuposing a prior representational system with the structure and resources of a natural language? Are the associated cognitive abilities constitutively linked to the language capacity? Are they necessary parts of the language acquisition device? These questions are addressed from the broader perspective of a general research program in comparative cognition, a program that has been carried out by Carey and Marc Hauser and whose results are reported in Carey's essay. In such studies the presence or absence in infant human cognition of given representational resources and abilities is compared with their presence or absence in nonhuman primate cognition. The underlying idea is that by identifying the cognitive features in question both from an ontogenetic and from an evolutionary standpoint one obtains a clearer grasp of their nature and of their connection with the language capacity. Thus, two questions have to be answered with a view to arriving at a satisfactory answer to the above questions. First, do prelinguistic human infants have concepts of object and number of the sorts under consideration? Second, are these representational resources available to nonhuman primates? Carey argues that evidence arising from research in comparative cognition indicates that both questions should be given affirmative answers, and hence that our concepts of object and number should be viewed as parts of a core of innate knowledge; such knowledge is claimed to be ontogenetically and evolutionarily prior to the system of symbolic representation embodied in our use and understanding of a natural language.

Carey takes the *object* concept as a sortal concept, a concept that provides us with criteria for individuating and counting the entities falling under it. The individuation principles associated with our concept of a physical object include: (a) principles about object permanence, particularly the idea that material objects are expected to continue to exist independently of us as time goes by and when perceptual contact with them is lost; (b) spatiotemporal principles, especially the idea that two material objects cannot occupy the same portion of space at the same time and the idea that there must be a continuous spatiotemporal path linking successive appearances of one and the same material object. These and other principles are said to be constitutive of our *object* concept. Drawing on empirical studies where the technique known as the *looking time* method is applied both to preverbal human infants and to nonhuman primates, Carey contends that creatures of both kinds can be credited with the concept of a physical object as defined by those features.

Contrary to Jean Piaget's account, on which the *object* concept is intrinsically linked to the language capacity, Carey takes research carried out by Elisabeth Spelke, Karen Wynn and others to show that the *object* concept, as well as the basic quantificational concepts *one*, *another*, are already available to prelinguistic human infants, namely babies with ages ranging from 2 and a half months to 3 months. Infants are described as knowing that material objects continue to exist when behind barriers and as making use of spatiotemporal principles such as the ones above for individuating material objects. A sharp contrast is drawn between the case of the sortal *object* and other object concepts, on the one hand, and the case of specific sortals, concepts such as *book*, *bottle*, and *doll*, on the other; indeed, Carey argues that human infants can only employ the latter range of concepts when the language device is already in place. Futhermore, in a series of studies undertaken by Carey and Hauser the looking time

technique was also applied to nonhuman primates, first to wild rhesus monkeys and then to cotton-top tamarins, creatures that are even more distant from us in evolutionary terms. The results obtained turn out to be analogous to those relative to infant cognition: nonhuman primates of both species are described as possessing the sortal *object* concept and the concepts of numerical identity and distinctness associated with the quantifiers *one, another*.

As to concepts of number, Carey discusses three proposals concerning the nature of the system used by human infants to represent the first three integers 1, 2, 3: the Numeron List Proposal, the Accumulator Proposal, and the Object File Proposal. She reads the literature in the area as providing us with evidence in support of the Object File Proposal, an account that distinguishes itself from the other two accounts in virtue of being definitely nonsymbolic; unlike them, the Object File Proposal is not based on the idea that a distinct symbol should be assigned to represent each integer. Carey concludes that "there is no evidence for a prelinguistic representational system of the same structure of natural language count sequences, such as '1, 2, 3…'" (p. 43). However, she is inclined to favour the Accumulator model with respect to the system employed by nonhuman primates to represent integers.

The issue of whether speechless creatures can or could have concepts and think is a highly controversial issue in the foundations of cognitive science. As we have seen, on Carey's view there is a range of basic concepts that are available to preverbal human infants and to nonhuman primates; it seems that such speechless creatures can be credited on that basis with thought and cognition.

A different and *prima facie* inconsistent approach can be found in Davidson's essay. Davidson investigates the requirements of thought, the conditions that an object – a system, a device, a creature – must in general satisfy in order to be correctly identified

or recognized as a cognising thing, in the sense of something that has concepts and is able to employ them in thought. He famously argues, here and elsewhere, that only systems or creatures that are very similar to us (mature human beings) can be legitimately credited with concepts and thoughts. One of the aspects that is taken by Davidson as crucial in establishing such similarity is precisely speech, mastery of a natural language; hence, on his view, speechless creatures in general and animals in particular do not literally have any concepts and are not literally capable of entertaining any thoughts. Of course, we may feel tempted to describe them as cognising things or thinkers; but such ascriptions are only projections we make on the basis of our own case, they should not be taken literally.

The gist of Davidson's main argument towards that conclusion is as follows. A necessary condition for a system or creature to possess concepts and exercise them in thinking is to have propositional attitudes, to have beliefs, desires, intentions, and so on. Propositional attitudes are something we ascribe by essentially using 'that'-clauses to specify propositional contents, e.g. as when we classify a creature as believing *that there is food in the fridge* or as wishing *that the rain stops*. Now, notwithstanding appearances to the contrary, animals and other speechless creatures are arguably incapable of having any propositional attitudes: no such attitude report can be literally true of them. Therefore, they lack concepts and thought.

In fact, Davidson invites us to equate having concepts and thinking with having propositional attitudes. Concepts are ways of sorting out items in the world, and to have a concept is to be able to, or to be disposed to, class or not a given item under the concept. But to be capable of recognizing an item as instantiating or not a concept is to be capable of judging or believing that it falls or does not fall under the concept; in other words, it is to be capable of having propositional attitudes. A difference should be

discerned between recognizing an *F* simpliciter – in the sense of being able to, or being disposed to, discriminate *F*s from non-*F*s – and recognizing an *F* as such (as an *F*); the latter ability but not the former requires a creature to be able to, or to be disposed to, judge or believe that the item in question is an *F*. Davidson dismisses in this way the temptation we very often have to ascribe concepts and thoughts to animals on the basis of their sometimes rather strong powers of discrimination, on the basis of their apparent ability to distinguish between items that do and items that do not fall under a concept; as he puts it, this ability consists in nothing other than mindless dispositions to respond in specific ways to items that *we describe* as instantiating the concept. Although it can surely be said of animals and other speechless creatures that they are able e.g. to *see* food in the fridge, it cannot be said of them that they are able to *see that* there is food in the fridge.

Davidson introduces three different but related sorts of requirement on having concepts or propositional attitudes; it turns out that each of those conditions can only be met by creatures endowed with a representational system with the structure and properties of a natural language. The first requirement is based on Davidson's holism about meaning and concept ascription. Concepts are individuated, not only in terms of certain relations they bear to items in the world, but also in terms of certain relations they bear to other concepts; some of the latter relations are entailments, others are relations of evidential support. Thus, one cannot possess a concept or have a belief without possessing a host of other concepts in a network of interrelated concepts or having a host of other beliefs in a web of beliefs. In order to be correctly classified as a thinker a creature must therefore be in possession of a relatively sophisticated conceptual repertoire, a system very much like the one embodied in a natural language. The second requirement is that in order to have concepts a creature must not only be in

a position to make occasional mistakes in applying a given concept, e.g. judging that a rat is crossing the street while watching a squirrel doing it, but also be in a position to recognize that the creature itself has made a mistake, which is clearly a higher-order propositional attitude. As Davidson puts it, "a creature that cannot entertain the thought that it may be wrong has no concepts, no thoughts" (p. 126); again, this involves crediting thinkers with sophisticated concepts such as the concepts of *error* and *objective truth*. Finally, a third set of requirements is imposed by concept possession upon the structure of thought and upon the structure of any language adequate to express it. Let us mention just the most basic of such conditions: (a) the creative property of concepts and thoughts, i.e. the fact that concepts can apply to an endless number of items and the fact that there are infinitely many thoughts to be entertained, requires that such a language contain demonstrative terms and truth-functional connectives; (b) assuming that one cannot have thoughts without a framework for objectual reference and without possessing the general concept of an object, the language should also contain resources adequate to play the role of variables and quantifiers in the usual symbolism of quantification theory (note the contrast between such a concept of an object and the *object* concept envisaged by Carey in her paper).

James Higginbotham's essay (chapter 10) focuses on semantics, the study of linguistic meaning, and its relation to the enterprise of cognitive science. Meaning is a key concept in the study of the mind/language connection. Indeed, an important feature of language processing, our ability to use and understand a natural language, consists in assigning meanings to spoken or written words and sentences of the language on particular occasions; and this is in some sense a mental activity, something that our minds do. How should word and sentence meaning be in general explained? How

should one characterize understanding, knowledge of meaning? Aspects of the twin topics of meaning and understanding are the subject of Higginbotham's paper.

Higginbotham approaches the topic of meaning from the standpoint of referential or truth-conditional semantics. According to this view, meaning is to be centrally explained in terms of reference and other related extensional notions, such as satisfaction and truth; all these are language-world relations, holding between linguistic expressions and items or sets of items in the world. Aspects of language use that apparently cannot be accounted solely in terms of referential properties, e.g. racial epithets and euphemism, are nevertheless only understood against the background of reference. On referential semantics, the meaning of a declarative sentence is given in terms of its truth-conditions, as these are compositionally determined on the basis of referential properties of lexical elements occurring in it and its syntax, the specific mode of combination of such elements in the sentence. The nuclear part of an account of meaning for a natural language takes the familiar Tarskian shape of an axiomatized body of statements, a *theory of meaning* for the language. The axioms of the theory specify referential properties for the primitives of the language, as well as the semantic significance of the available modes of combination; the theorems of the theory state truth-conditions for arbitrary sentences in the language according to their structure and the referential properties of their constituents as specified by the axioms.

The issue Higginbotham then addresses concerns the place that should be assigned to referential semantics within cognitive science. He assumes a familiar view of cognitive science, a view on which the mental states studied therein are computational in the following sense. They are individuated in part by their content or representational properties and in part by their causal powers (i.e. their interactions with other mental states and behaviour), the computations involved in the latter being

essentially symbolic, i.e. defined over a purely formal basis. The connection between referential semantics and cognitive science so construed is captured in the following two claims. First, understanding or knowledge of meaning consists in genuine propositional knowledge, although mostly tacit, possessed by articulate speakers and hearers of a natural language; it is *knowledge that* as contrasted with *knowledge how*. In other words, there is such a thing as semantic *competence* in Noam Chomsky's sense of the term: a system of internalised representations of semantic rules and principles. Second, the objects of such knowledge are conditions on reference as given by the statements of referential semantics; for instance, able speakers are said to have tacit knowledge of axioms stating the denotation of singular terms of the language. Knowledge of meaning is thus modelled as a complex computational mental state the input of which are the statements of referential semantics as embodied in a meaning theory as outlined above and the outputs of which are "behavior and adjustments of states that go to exemplify our rationality in the use of language" (p. 147).

These two claims play the role of major premises in the account of meaning and understanding proposed by Higginbotham. He argues indirectly in their support by examining a set of three alternative proposals that have been influential in semantic theory and by identifying their shortcomings when compared to referential semantics.

The first proposal derives from deflationary or minimalist accounts of reference, as well as of the other semantic notions (satisfaction, truth) that are central to referential semantics. A consequence of such views is that the notion of reference is inadequate to play any explanatory role in an account of linguistic meaning. At least in the version considered by Higginbotham, which he attributes to Paul Horwich, word and sentence meaning are rather explained in terms of concepts and propositions expressed, these being essentially extracted from the way we use words and sentences. According to

Higginbotham, the proposal founders because it is incapable of providing us with a satisfactory account of the significance of even such a simple sentential mode of combination as the predication schema; he argues that the proposal is unable to explain how we are able to know the meaning of a monadic predication on the basis of knowledge of the meanings of its components and appreciation of its syntactic structure. The second proposal rejected by Higginbotham is given in representationalist accounts of meaning and knowledge of meaning, views on which the objects of tacit knowledge are taken to be mental representations; he claims that such views turn out to be inconsistent with the nature of linguistic competence and even with the practice of linguistics itself. The third proposal is encapsulated in Jerry Fodor's semantic views, especially in his thesis that, strictly speaking, natural languages have no semantics; only mental representations, words and sentences in the language of thought, have a semantics (in fact, a referential semantics). Higginbotham discerns three independent ingredients in Fodor's views: (a) the thesis that to understand a sentence of a natural languager consists in mapping it onto a thought, the thought it conveys; (b) the thesis that such a mapping consists in translating the sentence into a sentence of the language of thought, a sentence which expresses the tought in question; and (c) the thesis that learning a natural language is learning how to map its sentences onto thoughts. Higginbotham argues that claims (a) and (b) are in the end consistent with referential semantics construed as a theory of knowledge of meaning; it is rather thesis (c), and the underlying idea that knowing a language is a practical ability, that should be rejected.

MIND AND WORLD

The dominant theme for a third and last group of essays is the connection between mind and world, a connection of lasting theoretical interest to researchers

working in cognitive science and its foundations. The mind/world connection is multifarious and can be decomposed into a number of different relations. It is convenient to use the familiar notion of a direction of fit to introduce some of these relations. One of them is knowledge, under which are subsumed several other central mind/world relations such as perception and memory; knowledge, as well as each of its species, can be described as having the mind-to-world direction of fit, in the sense that it is the mind that has to adapt to the world in order to produce knowledge. Aspects of knowledge are the subject of Zenon Pylyshyn's essay (chapter 12), in which issues involving visual perception are investigated, and of Christopher Peacocke's essay (chapter 11), in which issues about the integration of epistemology and metaphysics within a general theory of concepts are addressed. Another central relation between mind and world is action. Action can be described as having the world-to-mind direction of fit, in the sense that it is the world that adapts to the mind when it is acted upon by us and other creatures; issues involving the explanation of action and intelligent behaviour are discussed in John Searle's essay (chapter 13).

Holding at a more general level than knowledge, a mind/world relation that has always been at the heart of philosophical reflection, especially in philosophy of mind, is the relation known as Intentionality or Aboutness. (Action is also intentional, but in a different and more usual sense). Many mental states and events are said to be intentional in the sense of being about, or being directed upon, non-mental things; these are typically things in the world: specific objects, events, or situations. Take the mental state someone might be in when she believes that London is pretty; such a state is said to be intentional both in the sense of having an intentional object, the city of London itself, and in the sense of having an intentional content, that London is pretty. The same goes for a host of other mental states and events such as desires, fears, thoughts, regrets,

and so on. Intentionality cannot obviously be defined in terms of direction of it: knowledge is an example of an intentional state with the mind-to-world direction of fit and desire an example of an intentional state with the world-to-mind direction of it.

Aspects of the problem of intentionality or meaning (in one sense of 'meaning') – the problem of explaining how can there be things in the mind/brain, presumably worldly things, that are about other worldly things – are explicitly addressed in two papers in the present collection: Margaret Boden's essay (chapter 2) and Daniel Dennett's paper (chapter 9). In his essay Dennett launches a sustained criticism against the general picture of intentionality underlying two methodological claims that, on his view, have been omnipresent within the foundations of cognitive science. He labels these claims the *"content capture" assumption* and the *"isolated vehicles" assumption*. The first is a claim about content and (roughly) states that intentional content should be captured in terms of propositions or intensions; although Dennett has mostly in mind linguistic or symbolic representational resources, the claim is taken by him in a broader sense to mean that intentional contents should be specified in an explicit manner in terms of data structures, a term he employs to cover not only linguistic items like sentences in a language of thought but also representational devices like images, icons, and maps. The second claim is about the vehicles of intentional content and (roughly) states that it is possible, as well as theoretically desirable, to isolate such vehicles from the "outside" world; in particular, with respect to creatures like us (i.e. embodied nervous systems), there is an important level at which the bearers of intentionality – the *things-about-things* (as Dennett calls them) – can and should be taken in complete abstraction from their connections with the body.

According to Dennett, these two assumptions about intentionality have been rather pervasive in cognitive science. They owe their pervasiveness in part to their

original status as useful idealizations or oversimplifications; like any other science at its initial stages of development, cognitive science needs some such oversimplified assumptions in order to get off the ground. However, these assumptions have pave the way for a set of foundational questions about intentionality that are, on Dennett's view, the wrong questions to ask at the current stage of inquiry. One such question is the ontological issue about what sorts of things, what kinds of items in the mind or in the brain are the bearers of intentionality, the things-about-things. Are they mental representations, for instance sentences in the language of thought or else non-symbolic items like icons or maps, or are they just complex abilities generated by intricate structures (neural networks) in the brain? From the standpoint of a broadly evolutionary account of cognition and intentionality, an account on which intentional properties of mind or brain components are basically explained in terms of their evolutionary functions, Dennett argues towards the conclusion that those two enabling assumptions should be put aside as they give us a distorted picture of the intentionality present in embodied nervous systems; they should give way to alternative guiding assumptions, assumptions one may naturally extract from the evolutionary account.

Dennett claims that the "content capture" assumption leads us in the wrong direction and invites us to ask the wrong kind of questions. In a nutshell, his argument is that describing the intentional contents of given pieces of cognitive machinery in the mind/brain by specifying certain propositions or certain concepts, the propositions believed or the concepts entertained, does not give us in itself any idea of how these mind/brain components are able to carry out their intentional role; it does not give us any idea of how they are able to indicate, or contain information about, particular objects, properties, or events in the environment. In other words, content descriptions of such kind, the model of which is given in explicit expressions of intentional contents, do

not provide us with an adequate basis for an explanation of intentional properties; they are essentially incomplete in that respect. Intentional content should rather be approached in terms of functional descriptions, descriptions of the functions realized by the things-about-things in embodied nervous systems as such functions are determined by their evolution. Content specifications on the basis of propositional models are unable to do the work; at best, they give us only mnemonic labels for functional structures, labels that merely allude to intentional roles performed by things-about-things.

Appealing as it may be, the "isolated vehicles" assumption also conveys a misguided picture of the mind and its intentional properties. Dennett describes the cardinal idea behind the assumption by means of a suggestive image, an image which depicts the mind as a control system and the body as the system controlled by it. Just as the control systems of many machines can be isolated from the controlled systems, so that the latter systems can continue to operate even upon complete replacements of the former systems (e.g. remote controls of TV sets can be replaced without any loss of function); so the mind, although as material as the body, can be isolated from the body in the sense that its central activity as a processor of information is insensitive to the physical medium in which the information is transmitted, processed, and stored – in particular, it is not affected by the physics and chemistry of the body. Of course, there are points of contact between control system and controlled system, channels of information that link the mind to the body and to the "outside" world. These are of two kinds, transducer or input nodes and effector or output nodes, and their physical composition is surely relevant. Yet, according to the "isolated vehicles" assumption, it is possible to segregate transducers and effectors by regarding them as part of the controlled system, the body, not of the control system, the mind.

Dennett thinks this is in many respects a seductive picture. Nevertheless, the envisaged segregation is of doubtful coherence when applied to creatures like us and other embodied nervous systems.